

LISTING OF CLAIMS

1. (currently amended) An information processing method comprising:

providing an annotation for multiple page files, including the steps of:

obtaining a plurality of page files from a web site;
generating a group of said page files, page layout structures of which are at least similar by analyzing said page files to introduce structural descriptive forms for said page layout structures and characteristic values for said structural descriptive forms; employing said structural descriptive forms and said characteristic values to calculate an inter-page distance representing a similarity of said page files; and grouping said page files, of which said inter-page distance is equal to or smaller than a predetermined value;

providing a first annotation for an arbitrary page file in said group; and

correlating said first annotation with at least a part of other page files of said group.

2. (canceled)

3. (currently amended) The information processing method according to claim 1 ~~claim 2~~, wherein said structural descriptive forms are layout tags employing a style for designating a location on a page for representing tags that are correlated with said page layout structures included in said page files; and wherein said characteristic values are attributes of said layout tags and values of said attributes.

4. (currently amended) The information processing method according to claim 1 ~~claim 2~~, wherein said inter-page distance is obtained by calculating a sum of the values obtained by weighting said characteristic value and said structural descriptive form that is included in common with said multiple page files.

5. (original) The information processing method according to claim 1, wherein said step of correlating said first annotation with said other page files in said group includes the steps of:

determining whether said first annotation should be applied for the page files of said group;

adding a second annotation, when the determination is false, for an arbitrary page file of a page group consisting

of page files with which said first annotation is not correlated;

correlating said second annotation with at least a part of other page files of said page group; and

correcting a calculation expression for said inter-page distance, so that, at said step of generating a group, said page file with which said first annotation is correlated and said page files that are correlated with said second annotation do not fall in the same group.

6. (original) The information processing method according to claim 5, wherein said inter-page distance is calculated by using the sum of values obtained by weighting said characteristic value and said structural descriptive form that is included in common with said multiple page files; and wherein said calculation expression for said inter-page distance from a group of steps corrected by performing at least one step from a group of steps including:

an operation for increasing said weighting of said structural descriptive form and said characteristic value, for said structural descriptive form and said characteristic value that are different between said page file correlated with said first annotation and said page file correlated with said second annotation, and

an operation for reducing said weighting of said structural descriptive form and said characteristic value, for said structural descriptive form and said characteristic value that are in common with said page file correlated with said first annotation and said page file correlated with said second annotation.

7. (currently amended) The information processing method according to claim 1 ~~claim 2~~, further comprising the steps of:

introducing a representative structural descriptive form that represents said groups and a representative characteristic value for said representative structural descriptive form;

employing said representative structural descriptive form and said representative characteristic value to calculate an inter-group distance that delineates the similarity between said groups;

grouping said page files that are included in said groups, said inter-group distance of which is equal to or smaller than a predetermined value, and generating a common group;

adding an annotation to a common area wherein part of the page layout structure of an arbitrary file, included in

common for the members of said common group, is the same as or similar to at least a part of the page layout structure of a different page file; and

correlating said annotation with said common area provided for said different page file included, in common, for said common group.

8. (original) The information processing method according to claim 7, wherein said representative structural descriptive forms are layout tags employing a style for designating the location on a page for representing tags correlated with said page layout structures of said page files; and wherein said representative characteristic values are attributes of said layout tags and values of said attributes.

9. (original) The information processing method according to claim 7, wherein said inter-group distance is calculated by using the sum of the values obtained by weighting said representative characteristic value and said representative structural descriptive form that is included in common with said multiple groups.

10. (original) The information processing method according to claim 7, wherein said step of correlating said first annotation with said common area provided for said different page file includes the steps of:

determining whether said first annotation should be applied for said common area provided for the page files of said common group;

adding a second annotation, when the determination is false, to the common area of an arbitrary page file of a page group consisting of page files including said common area with which said first annotation is not correlated;

correlating said second annotation with 'Yes' part of the common areas of other page files of said page group; and

correcting a calculation expression for said inter-group distance, so that, at said step of generating a common group, said page file including said common area correlated with said first annotation and said page files including said common areas correlated with said second annotation do not fall in the same common group.

11. (currently amended) An information processing system, for providing an annotation for multiple page files, comprising:

means for obtaining page files from a web site;

means for generating a group of said page files, page layout structures of which are the same or similar
comprising means for analyzing said page files to introduce structural descriptive forms for said page layout structures and characteristic values for said structural descriptive forms; means for employing said structural descriptive forms and said characteristic values to calculate an inter-page distance representing the similarity of said page files; and means for grouping said page files, of which said inter-page distance is equal to or smaller than a predetermined value;

means for providing a first annotation for an arbitrary page file in said group; and

means for correlating said first annotation with 'Yes' a part of other page files of said group.

12. (canceled)

13. (currently amended) The information processing system according to claim 11 ~~claim 12~~, wherein said structural descriptive forms are layout tags employing a style for designating the location on a page for representing tags correlated with said page layout structures of said page files; and wherein said characteristic values are attributes of said layout tags and values of said attributes.

14. (currently amended) The information processing system according to claim 11 ~~claim 12~~, wherein said inter-page distance is calculated by using the sum of the values obtained by weighting said characteristic value and said structural descriptive form that is included in common with said multiple page files.

15. (currently amended) The information processing system according to claim 11 ~~claim 12~~, wherein said means for correlating said first annotation with said other page files in said group includes:

means for determining whether said first annotation should be applied for the page files of said group;

means for adding a second annotation, when the determination is false, for an arbitrary page file of a page group consisting of page files with which said first annotation is not correlated;

means for correlating said second annotation with 'Yes' part of other page files of said page group; and

means for correcting a calculation expression for said inter-page distance, so that, at said step of generating a group, said page file correlated with said first annotation

and said page files correlated with said second annotation do not fall in the same group.

16. (original) The information processing system according to claim 15, wherein said inter-page distance is calculated by using the sum of values obtained by weighting said characteristic value and said structural descriptive form that is included in common with said multiple page files; and wherein said calculation expression for said inter-page distance is corrected by performing at least one step from a group of steps including:

an operation for increasing said weighting of said structural descriptive form and said characteristic value, for said structural descriptive form and said characteristic value that are different between said page file correlated with said first annotation and said page file correlated with said second annotation, and

an operation for reducing said weighting of said structural descriptive form and said characteristic value, for said structural descriptive form and said characteristic value that are in common with said page file correlated with said first annotation and said page file correlated with said second annotation.

17. (currently amended) The information processing system according to claim 11 ~~claim 12~~, further comprising:

means for introducing a representative structural descriptive form that represents said groups and a representative characteristic value for said representative structural descriptive form;

means for employing said representative structural descriptive form and said representative characteristic value to calculate an inter-group distance that delineates the similarity between said groups;

means for grouping said page files that are included in said groups, said inter-group distance of which is equal to or smaller than a predetermined value, and generating a common group;

means for adding an annotation to a common area wherein part of the page layout structure of an arbitrary file, included in common for the members of said common group, is the same as or similar to at least a part of the page layout structure of a different page file; and

means for correlating said annotation with said common area provided for said different page file included in common for said common group.

18. (original) The information processing system according to claim 17, wherein said representative structural descriptive forms are layout tags employing a style for designating the location on a page for representing tags correlated with said page layout structures of said page files; and wherein said representative characteristic values are attributes of said layout tags and values of said attributes.

19. (original) The information processing system according to claim 17, wherein said inter-group distance is calculated by using the sum of the values obtained by weighting said representative characteristic value and said representative structural descriptive form that is included in common with said multiple groups.

20. (original) The information processing system according to claim 17, wherein said means for correlating said first annotation with said common area provided for said different page file includes:

means for determining whether said first annotation should be applied for said common area provided for the page files of said common group;

means for adding a second annotation, when the determination is false, to the common area of an arbitrary page file of a page group consisting of page files including said common area with which said first annotation is not correlated;

means for correlating said second annotation with 'Yes' part of the common areas of other page files of said page group; and

means for correcting a calculation expression for said inter-group distance, so that, at said means for generating a common group, said page file including said common area correlated with said first annotation and said page files including said common areas correlated with said second annotation do not fall in the same common group.

21. (original) An article of manufacture comprising a computer usable medium having computer readable program code means embodied therein for causing annotation, the computer readable program code means in said article of manufacture comprising computer readable program code means for causing a computer to effect the steps of claim 1.

22. (original) A program storage device readable by machine, tangibly embodying a program of instructions executable by the machine to perform method steps for annotation said method steps comprising the steps of claim 1.

23. (original) A computer program product comprising a computer usable medium having computer readable program code means embodied therein for causing annotation the computer readable program code means in said computer program product comprising computer readable program code means for causing a computer to effect the functions of claim 11.